

MitoBench: An interactive visual workbench for population genetics on mitochondrial DNA



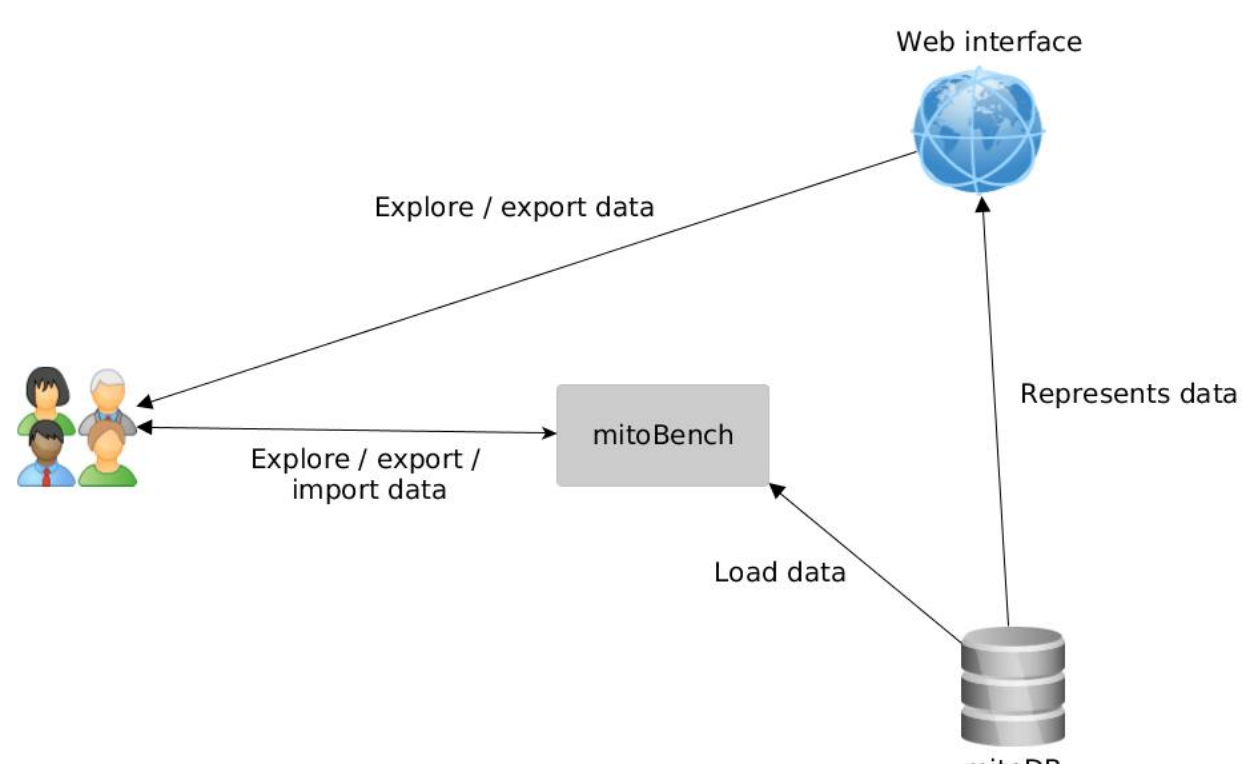
Judith Neukamm^{1,3*}, Alexander Peltzer^{1,2,3*}, Wolfgang Haak², Johannes Krause^{1,2} and Kay Nieselt³

1 Institute for Archaeological Sciences, Archaeo- and Paleogenetics, University of Tuebingen, Germany.
2 Max Planck Institute for the Science of Human History, Jena, Germany.
3 Integrative Transcriptomics, Center for Bioinformatics (ZBIT), University of Tuebingen, Germany.
*These authors contributed equally to the study.

Introduction

Despite the availability of modern next generation sequencing technologies and therefore nuclear human genomes, the sequencing and analysis of mitochondrial DNA (mtDNA) is still common. Especially in the research field of ancient DNA and the context of population genetics, mtDNA is often the only proxy available to study extinct populations and their relationship with modern populations. As a consequence, many population genetic studies rely on the analysis of mtDNA. A plethora of methods for the analysis of mtDNA exist, that address questions in population genetics, phylogeny and others. However, these tools typically rely on different file formats and often require manual interaction with the data for downstream analysis. Ultimately, these steps can be cumbersome, especially for non-bioinformaticians, resulting in an increased risk of errors during the analysis. To tackle these issues, we present MitoBench, a workbench to interactively analyze and visualize mitochondrial genomes with a focus on population genetics. The graphical user interface is kept simple, to accommodate even users without further prior knowledge on computational methods. Furthermore, it shows additional information such as metadata and statistics. Currently, MitoBench offers automatic file conversion tools to connect the workbench with existing analysis methods such as BEAST, Arlequin and others.

Workflow

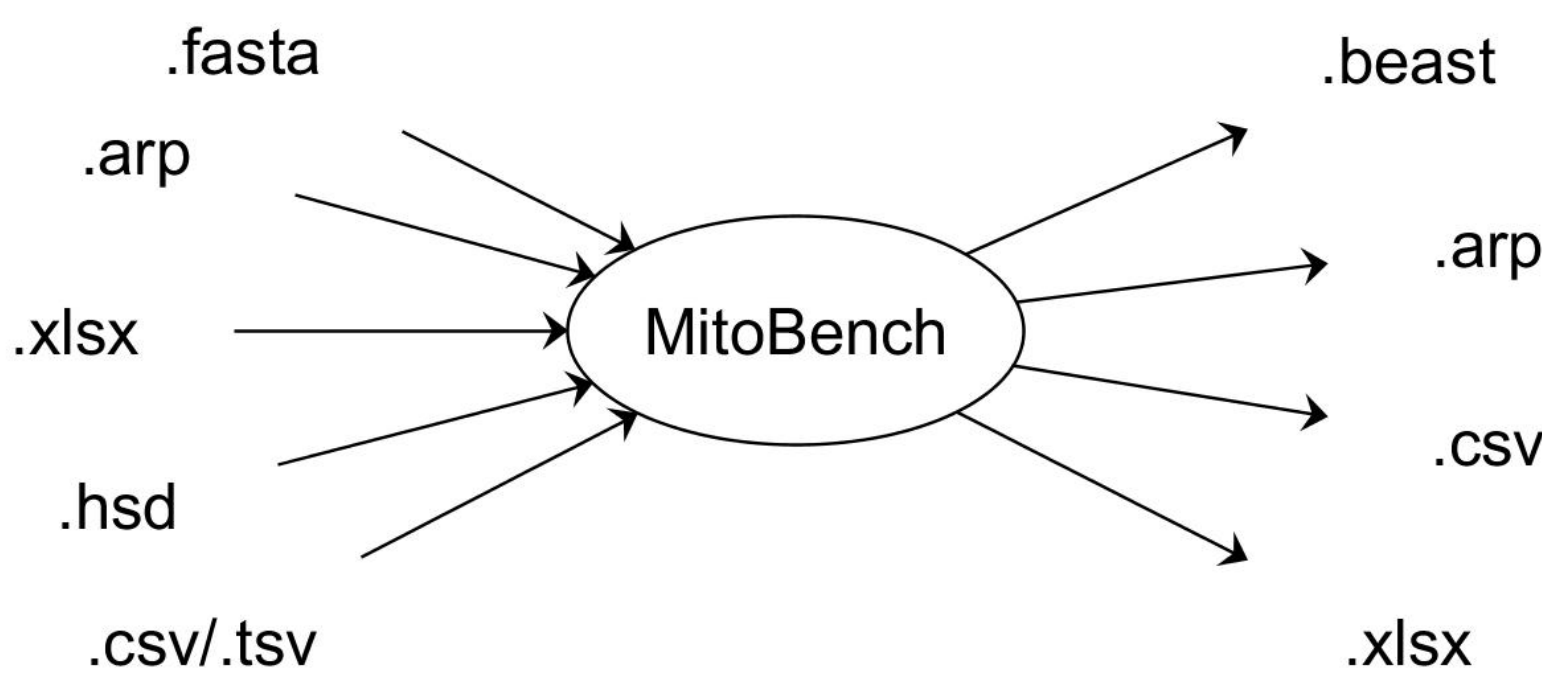


MitoDB

- Web-Frontend with Vaadin Java Framework
- Backend with PostgreSQL, providing sequence information and meta-data (SQL)
- Curated data upload with metadata
- Retrieval possible through WebUI and mitoBench

MitoBench

Data import and export



Data representation

User table								
ID	MTSequence	Haplogroup	Time Period	continent	country	country_region	culture_type	dating_lower
JC2880	GATCA...	T1a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-669
egypt_5AJ129	GATCA...	X1c	Egypt Modern	Africa	Egypt	Other	Modern	2015
egypt_6AJ98	GATCA...	H3w	Egypt Modern	Africa	Egypt	Other	Modern	2015
egpg3305866	GATCA...	L3F2a1	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egypt_6AJ63	GATCA...	R0a3	Egypt Modern	Africa	Egypt	Other	Modern	2015
egpg3305863	GATCA...	M1a5	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg3305862	GATCA...	R0a2	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg3306038	GATCA...	R0a2g	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg3306032	GATCA...	L0F2b	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg3306031	GATCA...	L32	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg3306030	GATCA...	L2a1d1	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egypt_7AJ139	GATCA...	I6b	Egypt Modern	Africa	Egypt	Other	Modern	2015
JC2879	GATCA...	U3b	Roman Period	Africa	Egypt	Beni Suef	Ancient	2015
JC2878	GATCA...	T1a7	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-235
egypt_BAL34	GATCA...	H20c	Egypt Modern	Africa	Egypt	Other	Modern	2015

Conclusions

MitoBench

- Easy file handling and file conversion
- Combination, manipulation and visualization of mitochondrial data
- Multiple filtering options
- Statistics

MitoDB

- Collection of mitochondrial data plus meta information
- Curated database

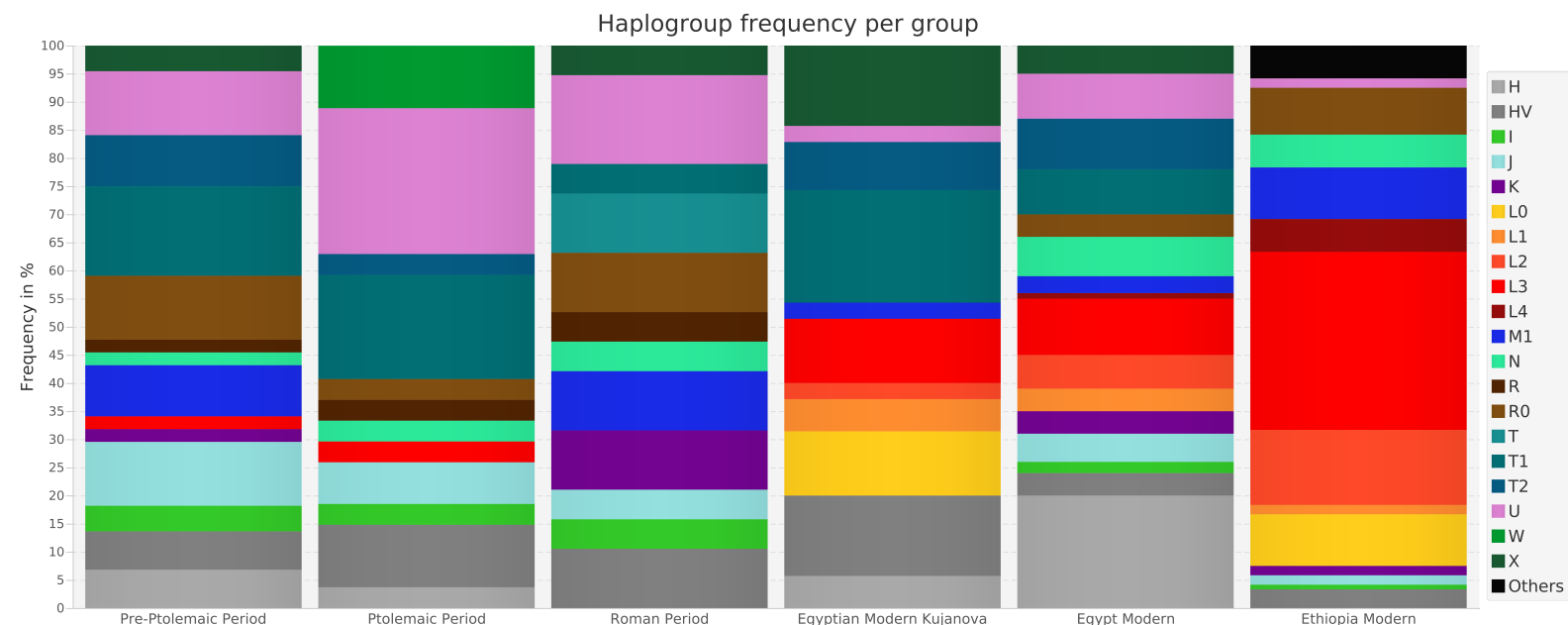
Data grouping

- interested in behaviour of different groups → Group data by feature
- internal grouping (columns not sorted)

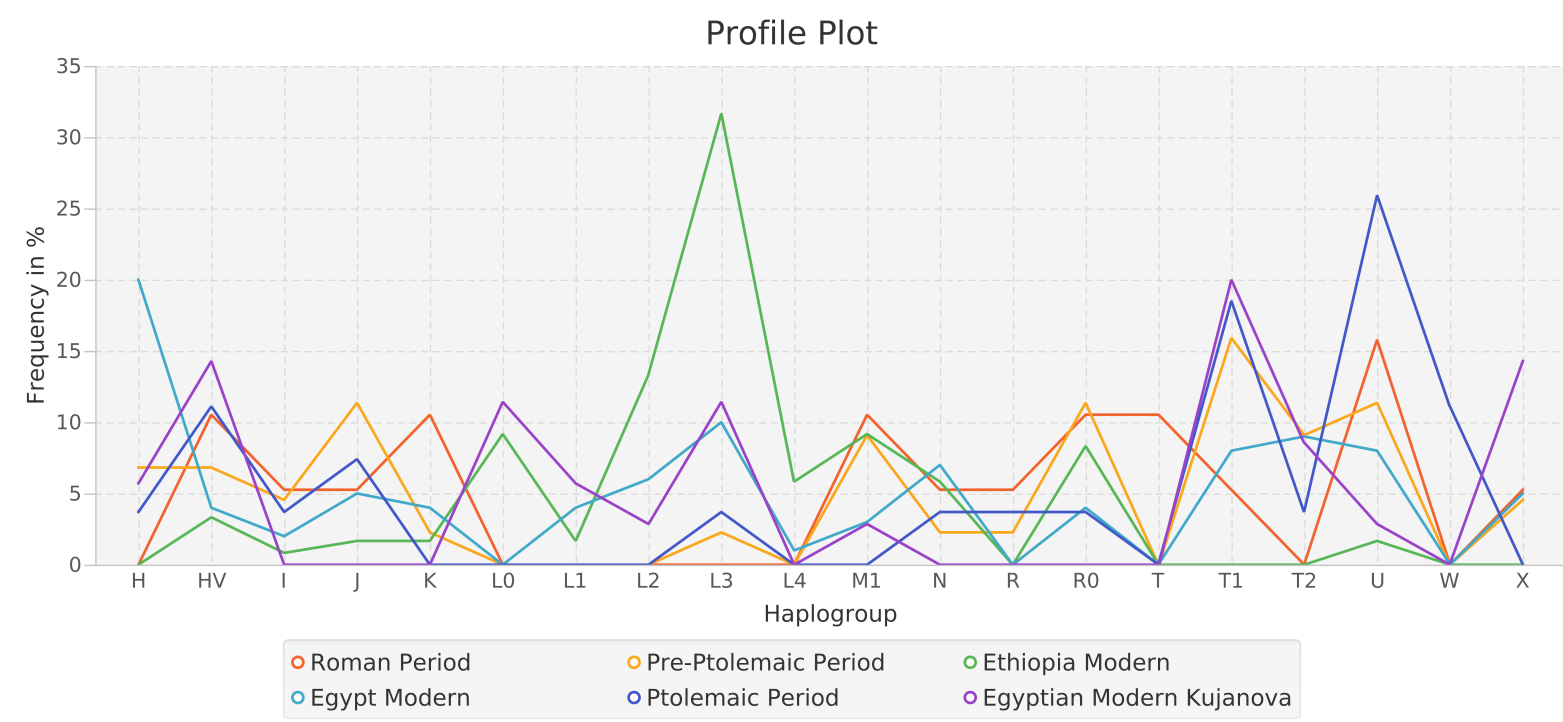
Data filtering / Statistics

- Haplogroup filtering
- Filtering based on Mutation
- Haplogroup frequencies
- Mutation frequencies

Data visualization



Data visualization



References

[1] Laurent Excoffier and Heidi EL Lischer. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Molecular ecology resources*, 10(3):564–567, 2010.

[2] Hansi Weissensteiner, Dominic Pacher, Anita Kloss-Brandstätter, Lukas Forer, Günther Specht, Hans-Jürgen Bandelt, Florian Kronenberg, Antonio Salas, and Sebastian Schönherr. Haplogrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic acids research*, page gkw233, 2016.

Contact

Judith Neukamm
University of Tübingen
Email judith.neukamm@uni-tuebingen.de
Phone +49 7071 2975651